# SED RDAS Frequently Asked Questions

What is the RDAS?

What SED data are available in the RDAS?

Do SED estimates produced in the RDAS match published SED Data Tables?

What are the benefits of using SED data in the RDAS?

What types of SED analysis can I create in the RDAS?

Can I create a trend analysis or compare different years of SED data in the same table?

What limits are in place for the number of variables that can be used in an analysis?

What limits are in place for filtering a table in the RDAS?

Why do the RDAS variables have "_R" at the end of their name? How are they different from the Doctorate Records File (DRF) versions?

Why are some of the RDAS tables suppressed?

How can I download the datasets available in the RDAS for use in another software package?

What are the footnotes at the bottom of my RDAS tables?

Where can I ask a question or send feedback on RDAS?

# SED RDAS Frequently Asked Questions

## What is the RDAS?

The Restricted Data Analysis System (RDAS) is an online tool designed to expand public access to the Survey of Earned Doctorates (SED) data. The SED RDAS data is based on a sample subjected to statistical procedures to produce estimates that reflect the entire population of doctorate recipients and protect the confidentiality of individual level data. The RDAS provides a secure platform to access a comprehensive set of SED variables and empowers users to create complex data tables at the click of a button.

## What SED data are available in the RDAS?

The RDAS provides information on the doctorate recipient's educational history, funding sources, postgraduation plans, and demographic characteristics. The RDAS currently has data for the 2018 and 2019 SED. For more information on the SED data items, see Survey of Earned Doctorates Survey Overview.

## Do SED estimates produced in the RDAS match published SED Data Tables?

Because the SED data in RDAS were subjected to sampling adjustment, missing data imputation, and statistical disclosure avoidance methods to protect disclosure of confidential information, SED estimates produced by RDAS may differ from the published SED data. See the SED RDAS Data Preparation Methodology for more information. See the SED RDAS Data Preparation Methodology for more information.

## What are the benefits of using SED data in the RDAS?

The missing data imputation adds value to SED data by utilizing responses from all survey responses and allowing users to create tables with estimates that fully represent the population. Furthermore, it makes the total number of doctorate recipients across different RDAS tables consistent.

## What types of SED analysis can I create in the RDAS?

The RDAS calculates averages, medians, percentage distributions, percent equal to a specified categorical value, percent not equal to a specified categorical value, percent greater than a specified categorical value, and percent less than a specified categorical value.

## Can I create a trend analysis or compare different years of SED data in the same table?

The RDAS does not currently offer trend analyses, but users may change the year selection in the top menu in the RDAS workspace to switch between 2018 and 2019 SED data. This functionality will be offered at a later date.

## What limits are in place for the number of variables that can be used in an analysis?

There are no limits to the number of column variables that can be used for average, median, percent equal to/not equal to, or percent greater/less than analyses, whereas percentage distribution calculations are limited to one column variable. There are no limits to the numbers of row variables that can be used. Filter variables are limited to a maximum of two discrete

variable filters per analysis. Nested row and nested column variables are limited to one variable per analysis. All analyses are limited to a maximum of three data table elements; therefore, users may only use three of the four available elements (i.e., column variable(s), row variable(s), nested variable, or filter variable) per analysis.

### What limits are in place for filtering a table in the RDAS?

Filtering is allowed, but there are restrictions on how many variables can be used in a single analysis. Filters can be used in combination with two other variables.

### Why do the RDAS variables have "_R" at the end of their name? How are they different from the Doctorate Records File (DRF) versions?

Variables with "_R" affixed to their name had their category values coarsened by combining categories with small counts with larger categories or by creating categorical variables from continuous variables. See the SED RDAS Data Preparation Methodology for more information.

### Why are some of the RDAS tables suppressed?

For individual disclosure protection, estimates are suppressed when there are too few cases within an analysis. The SED RDAS was designed to produce high-quality estimates while protecting against disclosure of confidential information.

### How can I download the datasets available in the RDAS for use in another software package?

Raw data are not available for download through the RDAS. SED microdata files may only be accessed under a restricted-use data license.

### What are the footnotes at the bottom of my RDAS tables?

The footnotes that appear at the bottom of analyses created in the RDAS contain additional information and source notes about the data in the table. Footnotes may include information such as details regarding special symbols displayed in the table or names of the RDAS variables used in the table, and a source note that indicates the survey year data used to generate the data table.

### Where can I ask a question or send feedback on RDAS?

Please contact the RDAS Help Desk at RDAS_support@rti.org with any questions or feedback.

# Statistical terms referenced in the RDAS

## Average
Calculated by summing the values in a category and dividing by how many respondents are in the category.

## Categorical variable
A variable that has at least two or more categories. Example: Marital status.

## Continuous variable
A variable that has an infinite number of possible values between any two values. Example: Salary.

## Confidence interval
The RDAS calculates 95% confidence intervals and is a range of numbers between which the true value falls between.

## Descriptive statistics
The descriptive statistics are provided for each variable within the variable information window. For categorical variables, the distribution of all values and their frequencies are reported. For continuous variables, the percent of positive values, the minimum value, the maximum value, the average, and the standard deviation are reported.

## Median
The midpoint of a frequency distribution or the middle number in an ordered list of numbers. The median separates the higher half from the lower half of the data sample.

## Percentage distribution
A frequency distribution where the total frequency is equated to one hundred and the individual class frequencies are expressed in proportion to that figure.

## Relative standard error (RSE)
The standard error of a survey estimate, divided by the estimate, and then multiplied by 100.

## Standard deviation
A measure of the amount of variation or dispersion in a set of values.

## Standard error
The standard deviation of the sampling distribution of a statistic.

## Variable name
A unique identifier for response items collected as part of the survey data collection. Variables or data elements are the individual items asked to survey respondents or items gathered as part of the data collection.